AD-A077 566 NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS F/6 12/1
ON ROBUST TESTS FOR HETEROSCEDASTICITY. (U)
JUL 79 R J CARROLL , D RUPPERT AFOSR-75-2796
UNCLASSIFIED MIMEO SER-1241 AFOSR-TR-79-1100 NL

1 OF 1
AD
A077566

END
DATE
FILMED
1 -80
DDC

AFOSR-TR- 79-1100

2

[redacted]

On Robust Tests for Heteroscedasticity

by

Raymond J. Carroll
and
David Ruppert

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1241

July 1979

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR-TR-79-1100
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): 6 On Robust Tests for Heteroscedasticity.

5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.

6. PERFORMING ORG. REPORT NUMBER: Mimeo Series No. 1241

7. AUTHOR(s): 10 Raymond J. Carroll [redacted] David Ruppert

8. CONTRACT OR GRANT NUMBER(s): 15 AFOSR-75-2796

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 16 61102F 2304 A5 17 A5

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research
Bolling AFB, DC 20332

12. REPORT DATE: 11 Jul[redacted]79

13. NUMBER OF PAGES: 11

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 13

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

14 MIMEO SER-1241

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Heteroscedasticity, robustness, linear model, M-estimates, regression, asymptotic tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We extend Bickel's (1978) tests for heteroscedasticity to include wider classes of test statistics and fitting methods. The test statistics include those based on Huber's function, while the fitting techniques include Huber's Proposal 2 (1977) for robust regression.

410064 Gm

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

On Robust Tests for Heteroscedasticity

by

Raymond J. Carroll* and David Ruppert**
University of North Carolina at Chapel Hill

We extend Bickel's (1978) tests for heteroscedasticity to include wider classes of test statistics and fitting methods. The test statistics include those based on Huber's function, while the fitting techniques include Huber's Proposal 2 (1977) for robust regression.

AMS 1970 Subject Classifications: Primary 62G35; Secondary 62J05, 62J10.

Key Words and Phrases: Heteroscedasticity, robustness, linear model, M-estimates, regression, asymptotic tests.

*This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-75-2796.

**This work was supported by the National Science Foundation Grant NSF MCS78-01240.

1. Introduction

We consider the general linear model

$$(1.1) \qquad Y_i = \tau_i + \sigma(\tau_i,\theta)\varepsilon_i, \quad \tau_i = \underline{c}_i' \underline{\beta}_0 \quad (i = 1,\ldots,n),$$

where $\underline{\beta}_0$ is an unknown $(p \times 1)$ vector, the $(p \times 1)$ vectors $\underline{c}_i'$ are known, the error terms ε_i are independent and identically distributed (i.i.d.) with common distribution function F, and $\sigma(\tau_i,\theta)$ expresses the possible heteroscedasticity in the model, with

$$(1.2) \qquad \sigma(\tau,\theta) = 1 + \theta\, a(\tau) + o(\theta) \quad \text{as } \theta \to 0 .$$

Bickel (1978), generalizing work of Anscombe (1961), defines robust tests for heteroscedasticity, which in the present context is a test of H_0: $\theta = 0$; the idea is to replace aspects of the usual informal examination of residuals by formal statistical inference about the probability structure of the data. If $\{t_i\}$ are the fitted values (from least squares or possibly a robust regression method (Huber (1973)(1977))) and b is an even function, Bickel's robust test statistic is

$$(1.3) \qquad A_b = \sum_{i=1}^{n} (a(t_i) - a_1(t)) b(r_i)/\hat{\sigma}_b ,$$

where

$$(1.4) \qquad r_i = Y_i - t_i = \text{residual},$$

$$(1.5) \qquad \hat{\sigma}_b^2 = \sum_{i=1}^{n} (a(t_i) - a.(t))^2 \,(n-p)^{-1} \sum_{i=1}^{n} (b(r_i) - b.(r))^2 ,$$

and for any function g,

$$g.(x) = n^{-1} \sum_{i=1}^{n} g(x_i) .$$

Bickel makes the following assumption:

(1.6) b is bounded and has two continuous, bounded derivatives.

Under (1.6) and other assumptions (see Theorem 1 below), Bickel obtains the asymptotic distribution of A_b under H_0: $\theta = 0$ and contiguous alternatives; results are obtained for the case $p^2/n \to 0$.

One of the most attractive choices of b (well motivated in Bickel's Section 3) is Huber's function squared:

$$b(x) = x^2 \qquad |x| \leq k$$
$$= k^2 \qquad |x| > k \,. \tag{1.7}$$

This choice of b does not satisfy (1.6) so that Bickel's Theorem 3.1 does not apply. He states that the strong smoothness condition (1.6) is "unsatisfactory" and obtains results for (1.7) only when p is bounded and fitting is by least squares.

In this note we show by a simple modification of Bickel's proofs (using techniques of Carroll (1978)), that results for A_b can be obtained for b given by (1.7) even when $p^4/n \to 0$ and fitting is by robust estimates on least squares. This result is given in Section 2. In Section 3 we note extensions which obtain scale invariance by robust estimation with scale estimated by Huber's Proposal 2.

2. Main Results

Where possible we adopt Bickel's notation. Without loss of generality we assume $n^{-1} \Sigma \underline{c}_i' \underline{c}_i = I$. To provide a frame of reference we state:

Theorem 1 (Bickel (1978)). Suppose the following hold:

$$\max |\tau_i| \leq M \,. \tag{2.1}$$

(2.2) $$n^{-1} \Sigma(a(\tau_i) - a_.(\tau))^2 \geq M^{-1} > 0,$$

(2.3) $$|\theta n^{\frac{1}{2}}| \leq M,$$

(2.4) F is symmetric about zero,

(2.5) $M^{-1} \leq J_2(F) \leq M$ where for $F' = f$ (absolutely continuous),

$$J_2(F) = \int (x f'(x)/f(x) + 1)^2 f(x)dx,$$

(2.6) $$\mathrm{Var}(b(\varepsilon_1)) \geq M^{-1} > 0 ,$$

(2.7) The function a is twice boundedly and continuously differentiable.

(2.8) $$\text{If } d_i = t_i - \tau_i,\ \Sigma d_i^2 = O_p(p),$$

(2.9) $$b(x) = b(-x),$$

(2.10) b is bounded and satisfies (1.6),

(2.11) $$p^2/n \to 0 .$$

Then

(2.12) $$P_\theta\{\Lambda_b \geq z\} = 1 - \Phi(z - \Lambda_b) + o(1) ,$$

where

(2.13) $$\Lambda_b(\theta,n) = \theta[\sum_{i=1}^{n} (a(\tau_i) - a_.(\tau))^2]^{\frac{1}{2}} E\, \varepsilon_1 b'(\varepsilon_1)[\mathrm{Var}(b(\varepsilon_1))]^{-\frac{1}{2}} .$$

Our generalization of Theorem 1 to incorporate such functions as (1.7) is

Theorem 2. Suppose (2.1)-(2.9) and the following hold:

(2.14) b is bounded, Lipschitz of order one, and has two bounded continuous derivatives except possibly at a finite number of points, which we take as ±c.

$$p^4/n \to 0 . \tag{2.15}$$

Then (2.12) holds. (Assumption (2.8) is discussed in the next section.)

<u>Proof of Theorem 2.</u> The key results in Bickel's proof are (A34)-(A37) with

$$\begin{aligned} w_{ij} &= 1 - 1/n \qquad (i = j) \\ &= -1/n \qquad (i \neq j) . \end{aligned}$$

Because b is bounded and Lipschitz of order one, (A34)-(A36) follow exactly as given by Bickel. He uses (A37) to prove

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^{n} (a(t_i) - a_.(t))b(r_i) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (a(\tau_i) - a_.(\tau))b(\varepsilon_i) \\ &+ n^{-\frac{1}{2}} E\, b'(\varepsilon_1) \sum_{i=1}^{n} (a(\tau_i) - a_.(\tau))d_i + o_p(1) , \end{aligned} \tag{2.16}$$

where $d_i = t_i - \tau_i$. Instead of proving (A37) we will prove (2.16) directly. As seen in Bickel's (A41)-(A47), (2.16) is verified by proving either (A48) (as Bickel has done) or

$$n^{-\frac{1}{2}} A_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\, a(\tau_i)(b(r_j) - b(\varepsilon_j) + d_j\, b'(\varepsilon_j)) \xrightarrow{p} 0 . \tag{2.17}$$

We will prove (2.17). Note that in Bickel's proofs of (A41)-(A47) the assumption (1.6) is not needed; the weaker assumption (2.14) suffices. Since $r_j = \varepsilon_j - d_j$, with I being the indicator function, rewrite

$$\begin{aligned} A_n &= \Sigma_i\, \Sigma_j\, w_{ij}\, a(\tau_i)(b(\varepsilon_j - d_j) - b(\varepsilon_j) + d_j\, b'(\varepsilon_j)) \\ &\times \{I(-c + a_n \le \varepsilon_j \le c - a_n) + I(c - a_n \le \varepsilon_j \le c + a_n) \\ &+ I(-c - a_n \le \varepsilon_j \le -c + a_n) + I(\varepsilon_j \ge c + a_n) + I(\varepsilon_j \le -c - a_n)\} \\ &= A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5} , \end{aligned} \tag{2.18}$$

where $a_n \to 0$ will be specified later. We also write $A_n = \Sigma_i \Sigma_j K_{ij}(n)$. We can further write

$$(2.19)\quad A_{n1} = \Sigma_i \Sigma_j K_{ij}(n)\left\{\begin{matrix} I(-c < \varepsilon_j - d_j < c \\ + I(|\varepsilon_j - d_j| > c)\end{matrix}\right\} I(-c+a_n<\varepsilon_j<c-a_n) = A_{n1}^{(1)} + A_{n2}^{(2)}$$

As in Bickel's (A48), since b is differentiable on $(-c,c)$,

$$|A_{n1}^{(1)}| = O_p(\Sigma d_i^2) = O_p(p) .$$

Note that if $-c + a_n < \varepsilon_j < c - a_n$ and $|\varepsilon_j - d_j| > c$ then $|d_j| > a_n$. Then, by (2.1), since b is Lipschitz and $w_{ij} = \delta_{ij} - 1/n$,

$$|A_{n1}^{(2)}| \le M_1 \sum_{j=1}^{n} |b(\varepsilon_j - d_j) - b(\varepsilon_j) + d_j b'(\varepsilon_j)| I\{-c + a_n < \varepsilon_j < c - a_n, |\varepsilon_j - d_j| \ge c\}$$

$$\le M_2 \sum_{i=1}^{n} |d_j| I\{|d_j| \ge a_n\} \le M_2 (\sum_{j=1}^{n} d_j^2)^{\frac{1}{2}} (\sum_{j=1}^{n} I\{|d_j| > a_n\})^{\frac{1}{2}} = O_p(p/a_n) .$$

Thus $|A_{n1}| = O_p(p/a_n)$. Similarly, $|A_{n4}| = O_p(p/a_n)$, $|A_{n5}| = O_p(p/a_n)$.

Further, by (2.1) and since *b* is *Lipschitz*,

$$(2.20)\qquad |A_{n2}| \le M_1 (\Sigma d_j^2)^{\frac{1}{2}} (\Sigma_j I\{c - a_n \le \varepsilon_j \le c + a_n\})^{\frac{1}{2}} .$$

A similar bound holds for $|A_{n3}|$. Since by (2.5) F is Lipschitz in neighborhoods of $\pm c$, Lemma 1 of Carroll (1978) shows

$$(2.21)\qquad \sum_{j=1}^{n} I\{c - a_n \le \varepsilon_j \le c + a_n\} = O_p(\max(\log n, n(F(c + a_n) - F(c - a_n))) .$$

Provided $na_n \ge \log n$, (2.20) and (2.21) give

$$|A_{n2}| = O_p((n\ pa_n)^{\frac{1}{2}}) \qquad |A_{n1}| = O_p((n\ pa_n)^{\frac{1}{2}}) .$$

This yields

$$(2.22)\qquad n^{-\frac{1}{2}} A_n = O_p((pa_n)^{\frac{1}{2}} + n^{-\frac{1}{2}} (p/a_n)) .$$

If we take $a_n = n^{-\frac{1}{4}}$, (2.15) and (2.22) yield

$$n^{-\frac{1}{2}} A_n = o_p(1) ,$$

completing the proof. □

3. Extensions

A. Scale invariance

The test statistic A_b is not scale invariant. To obtain such invariance, one would rewrite the model (1.1)-(1.2) so that $\sigma(\tau,\theta) = (1 + a(\tau)\theta + c(\theta))/\sigma_0$, where σ_0 is a scale parameter consistently estimated (when $\theta = 0$) by a scale estimate $\hat{\sigma}$ (provided by least squares or Huber's Proposal 2 for robust regression). To obtain scale invariance, Bickel suggests replacing $b(r_i)$ by $b(r_i/\hat{\sigma})$. The statements and proofs of Theorems 1 and 2 must be modified for this new test statistic which we denote $A_b(\hat{\sigma})$. An analogue of Theorem 2 is

Theorem 3. Suppose the conditions of Theorem 2 hold and, in addition,

(3.1) $$n^{\frac{1}{2}} (\hat{\sigma} - \sigma_0) = O_p(1) ,$$

(3.2) $$E\{b'(\varepsilon_1)\varepsilon_1\}^2 < \infty ,$$

(3.3) $$E\{b''(\varepsilon_1)\varepsilon_1\}^2 < \infty .$$

Then (2.12) holds for $A_b(\hat{\sigma})$.

Remark. Assumption (3.1) is the subject of part B of this section. Assumptions (3.2) and (3.3) hold if b is constant outside an interval (as is (1.7)).

Sketch of the proof of Theorem 3. We need to verify substitutes for Bickel's (A35) and (A37) when $b(r_i)$ is replaced by $b(r_i/\hat{\sigma})$, $b(\varepsilon_i)$ is replaced by $b(\varepsilon_i/\sigma_0)$ and the remainder terms are (respectively) $O_p((np)^{\frac{1}{2}})$ and $O_p(p)$. To prove the substitute for (A35) one must show that

$$(3.4)\quad \Sigma\, w_{ij}\, b(r_i/\hat{\sigma})\, b(r_j/\hat{\sigma}) - \Sigma\, w_{ij}\, b(\varepsilon_i/\hat{\sigma})\, b(\varepsilon_j/\hat{\sigma}) = O_p((np)^{\frac{1}{2}})$$

$$(3.5)\quad \Sigma\, w_{ij}\, b(\varepsilon_i/\hat{\sigma})\, b(\varepsilon_j/\hat{\sigma}) - \Sigma\, w_{ij}\, b(\varepsilon_i/\sigma_0)\, b(\varepsilon_j/\sigma_0) = O_p((np)^{\frac{1}{2}}) .$$

Using the special form of w_{ij}, (3.4) follows from (2.8) and the fact that b is bounded and Lipschitz; (3.5) is a consequence of (3.1) and (3.2). Statement (A37) is more complex. The analogue of (A42)-(A45) is to show

$$\sum_{i,j} w_{ij}(a(\tau_i) - a(t_i))\, b(\varepsilon_j/\hat{\sigma}) = O_p(p) ,$$

for which (using Bickel's proof) it suffices to show

$$(3.6)\qquad \sum_{i,j} w_{ij}\, a'(t_i)\, d_i(b(\varepsilon_j/\hat{\sigma}) - b(\varepsilon_j/\sigma_0)) = O_p(p) .$$

We rewrite (3.6) as

$$\begin{aligned}(3.7)\qquad & \sum_{i,j} w_{ij}\, a'(\tau_i) d_i\, (\varepsilon_j\, b'(\varepsilon_j/\sigma_0) - E\,\varepsilon_1\, b'(\varepsilon_1/\sigma_0)(1/\hat{\sigma} - 1/\sigma_0) \\ & + \sum_{i,j} w_{ij}\, a'(\tau_i) d_i\, (b(\varepsilon_j/\hat{\sigma}) - b(\varepsilon_j/\sigma_0) - (1/\hat{\sigma} - 1/\sigma_0)\varepsilon_j\, b'(\varepsilon_j/\sigma_0) \\ & = B_{1n} + B_{2n} .\end{aligned}$$

That $B_{1n} = O_p(p)$ follows from using the Schwarz inequality, the boundedness of a, and then applying (3.1) and (3.2). That $B_{2n} = O_p(p)$ is complicated notationally but is a consequence of a weakened version of Lemma 2 of Carroll (1978). This verifies (3.6).

The analogue of (A48) is to show that

$$(3.8)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)(b(r_j/\hat{\sigma}) - b(\varepsilon_j/\sigma_0)) = O_p(p) .$$

First note that (3.1) and the proof of Theorem 2 can be used to show that

$$(3.9)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)(b(r_j/\hat\sigma) - b(\varepsilon_j/\hat\sigma) + (d_i/\hat\sigma)b'(\varepsilon_j/\hat\sigma)) = O_p(p)\ .$$

To verify (3.8) we must show that the difference between (3.8) and (3.9) is $O_p(p)$; this is a consequence of the following:

$$(3.10)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)(b(\varepsilon_j/\hat\sigma) - b(\varepsilon_j/\sigma_0)) = O_p(p)$$

$$(3.11)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)\,\frac{d_j}{\hat\sigma}\,(b'(\varepsilon_j/\hat\sigma) - b'(\varepsilon_j/\sigma_0)) = O_p(p)$$

$$(3.12)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)\, d_j\, b'(\varepsilon_j/\sigma_0)(1/\hat\sigma - 1/\sigma_0) = O_p(p)\ .$$

Equations (3.11) and (3.12) follow by applying the Schwarz inequality, (2.8), (2.14), (3.1) and (3.3). We can rewrite (3.10) as

$$(3.13)\qquad \sum_{i,j} w_{ij}\, a(\tau_i)(b(\varepsilon_j/\hat\sigma) - b(\varepsilon_j/\sigma_0) - (1/\hat\sigma - 1/\sigma_0)\varepsilon_j\, b'(\varepsilon_j/\sigma_0))$$

$$+ \sum_{ij} w_{ij}\, a(\tau_i)[\varepsilon_j\, b'(\varepsilon_j/\sigma_0) - E\,\varepsilon_1\, b'(\varepsilon_1/\sigma_0)](1/\hat\sigma - 1/\sigma_0) = B^*_{n1} + B^*_{n2}\ ,$$

the last step following since $\Sigma\, w_{ij}\, a(\tau_i) = 0$. That $B^*_{n1} = O_p(p)$ follows as in the proof of Theorem 2, while $B^*_{n2} = O_p(p)$ follows from (3.1) and the Chebychev inequality. □

B. <u>On assumption (2.8)</u>. Huber's Proposal 2 for robust regression is to solve

$$(3.12)\qquad \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i\,\underline{\beta})/\sigma)\underline{c}_i = 0$$

$$(3.13)\qquad \sum_{i=1}^{n} \psi^2((Y_i - \underline{c}_i\,\underline{\beta})/\sigma) = \zeta = E\,\psi^2(Z)\ ,$$

the last expectation taken under the standard normal distribution function. Huber (1973) shows (2.8) under the following conditions:

(3.14) ψ is odd and non-decreasing,

ψ has two bounded continuous derivatives,

$\sigma = 1$ and only (3.12) is solved,

$\gamma p \to 0$, where γ is the maximum diagonal element of the projection matrix $C(C'C)^{-1} C' = CC' = \Gamma$ (since $\gamma \geq p/n$, $p^2/n \to 0$ is necessary.)

The above conditions are very restrictive in assuming σ known, and Huber's function ($\psi(x) = \max(-k, \min(k,x))$) does not satisfy the smoothness condition.

Carroll and Ruppert (1979) have generalized Huber's result by means of the techniques used in the proof of Theorem 2. They utilize the full system (3.12)-(3.13) and assume only that ψ satisfies (2.14), which is true for all functions used in practice. The price paid is a stronger condition on the growth rate of p; they require that for some sequence $a_n \to 0$, both $\gamma p/a_n^2 \to 0$ and $n\gamma a_n \to 0$. When the design is balanced ($\gamma = p/n$), then $a_n = p^{-(1+\varepsilon)}$ for any $\varepsilon > 0$ suffices, but that requires $p^{4+2\varepsilon}/n \to 0$.

C. Smoothness of F. Condition (2.5) is rather strong. Ruppert and Carroll (1979) show by entirely different methods that when p is fixed and b satisfies (2.14), (2.5) can be relaxed by requiring only that F is Lipschitz of order one in neighborhoods of $\pm c$.

References

Anscombe, F.J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1*, 1-36

Bickel, P.J. (1978). Using residuals robustly I: tests for heteroscedasticity, non-linearity. *Ann. Statist. 6*, 266-291.

Carroll, R.J. (1978). On almost sure expansions for M-estimates. *Ann. Statist.* 6, 314-318.

Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte-Carlo. *Ann. Statist. 1*, 799-821.

Huber, P.J. (1977). Robust Statistical Procedures. SIAM, Philadelphia.